Group Report 1964-17

Spectra of Vocoder Channel Signals

C. M. Rader

4 May 1964

Prepared under Electronic Systems Division Contract AF 19(628)-500 by

Lincoln Laboratory
MASSACHUSETTS INSTITUTE OF TECHNOLOGY
Lexington, Massachusetts

The work reported in this document was performed at Lincoln Laboratory, a center for research operated by Massachusetts Institute of Technology, with the support of the U.S. Air Force under Contract AF 19(628)-500.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

LINCOLN LABORATORY

SPECTRA OF VOCODER CHANNEL SIGNALS

C. M. RADER

Group 62

GROUP REPORT 1964-17

4 MAY 1964

LEXINGTON MASSACHUSETTS

ABSTRACT

A channel vocoder derives the short time magnitude spectrum of speech by passing the speech through a set of analyzing channels, each consisting of a band pass filter followed by an envelope detector and a low pass filter. Each channel produces a slowly varying output signal which represents a short term average of the speech energy in the corresponding frequency band. The measurement of the power spectra of these signals is described herein. It appears that these signals may be thought of as band limited to 25 cps.

Accepted for the Air Force
Franklin C. Hudson, Deputy Chief
Air Force Lincoln Laboratory Office

I. INTRODUCTION

The analyzer section of a vocoder produces a band-compressed version of speech by deriving from the speech a set of slowly varying signals. One of these slowly varying signals is the excitation or pitch. The other signals are the envelopes of narrow-band portions of the speech signal. The properties of these envelopes are of importance to vocoder designers because much of the bandwidth (and equipment) of a vocoder is devoted to adequate transmission of such envelopes.

Most of the vocoder systems built to date have employed a low pass filter following the envelope detector of each analyzing channel. The low pass filter eliminates the high frequencies (harmonics of the pitch) from the channel output. It can also smooth the envelope. It is not known to what extent smoothing of the envelope is related to degradation of the synthesized speech, but typical vocoder low pass filters cut off at several tens of cycles per second. The envelopes resulting from speech inputs vary slowly compared to such cut off frequencies. More general channel inputs would produce output envelopes varying at a rate determined by the channel bandwidth, which is at least 120 cps in the channels of the Lincoln vocoder.

This report describes the measurement of the power spectra of the envelopes in several channels of a vocoder to aid in the determination of:

1. the requirements on the low pass filter and
2. an adequate sampling rate for digital transmission of the envelopes.

II. MEASUREMENT TECHNIQUE

The operation of a vocoder has been described in a previous Group Report.* We shall be concerned with the envelope in one of the spectrum channels of the analyzing portion of the vocoder. Figure 1 shows such a channel. The output of the detector has a component near dc, representing the envelope of the band pass filtered speech, and several higher frequency components. The low pass filter passes only the envelope, which is the desired channel output. This envelope shall be referred to as a channel signal.

Typical vocoders have from ten to forty channels. In the Lincoln vocoder the detectors and low pass filters are identical in all the channels. The only circuit that differs from channel to channel is the band pass filter; a channel will usually be identified by stating the pass band of the filter.

With speech as an input, most of the energy of a channel signal lies below 30 cps. This has greatly hampered previous spectral measurements, because it is difficult to record low frequency signals, and because it is expensive to construct, for use in a spectrograph, band pass filters which are narrow enough to give useful spectral resolution of signals occupying only tens of cps.

Holmes† has modulated a 270 cps square wave with a channel signal, recorded it on magnetic tape at 1 7/8 inches per second, and played it back at 15 inches per second. He then performed a spectral analysis of the resulting

---

* B. Gold, "Vocoded Speech," 62G-1, 7 May 1963.

† J.N. Holmes, "Some Measurements of the Spectra of Vocoder Channel Signals," Research Report No. 20651, (British) Post Office Research Station, Dollis Hill, London, (April 24, 1961).

signal with a conventional sound spectrograph (whose narrowest analyzing filter was 15 cps wide) and deduced the spectrum of the channel signal from this analysis. Even with such a technique, the frequency resolution is only about 2 cps, and the modulation and demodulation devices, as well as the sound spectrograph itself, limit the dynamic range of the measurements.

For these reasons it was decided to compute the spectrum of a channel signal on the TX-2 digital computer. This method promised the advantages of:

1. overcoming the problem of dc recording,
2. allowing as fine a spectral resolution as desired,
3. requiring no new equipment to be built, and
4. simplifying the future processing and presentation of the data.

Two limitations are:

1. the requirement of considerable computer time to perform spectral analysis, and
2. the error introduced by sampling and quantizing the channel signal for storage in the computer memory.

The second limitation is quite negligible for this particular problem. A considerable effort was made to minimize the computation time required.

The channel signals used in the experiment were generated for the computer by playing tape-recorded speech through an analyzing channel identical to those in the Lincoln vocoder. (See Fig. 2.) The signals were sampled every millisecond and quantized into 256 levels by the computer's analog-to-digital converter. (After quantization, each signal was examined to make sure that most but not all of the 256 levels were occupied. If not, the amplitude scale was changed and the process repeated.) The use of 256 levels was dictated by the convenience of rapid multiplication of eight bit positive numbers on TX-2, and seems justified on the basis of experience

with digitized vocoders, most of which use fewer levels, logarithmically spaced.

Two thousand quantized samples generated from an input sentence two seconds long were stored in the computer memory for processing by a spectral analysis program. The two-second sentence duration permitted a theoretical spectral resolution to one half cps.

The low pass filter used in the channel, designed for maximally-flat envelope delay, was a fifth order Bessel type. The measured frequency response of the filter in the range of interest, from dc to 32 cps is shown in Figure 3. Results presented here may therefore be corrected by subtracting the attenuation shown in Figure 3 if the results are to be interpreted as properties of speech. They should not be so corrected, however, when used to evaluate a transmission scheme for the Lincoln vocoder.

Twenty input sentences were used in the experiment, five each spoken by two male and two female speakers; each was approximately two seconds long. The same sentences were processed by eighteen different channels. The channels used, listed in Table 1, included all the 120 cps wide channels below 1380 cps, five of the channels between 1500 and 3060 cps, and two wider channels (360 cps) above 3060 cps. The frequency and impulse responses shown in Figure 4 are from a typical band pass filter used in the experiment.

Since the time of these measurements several improvements have been made in the vocoder band-pass filters. The impulse responses of the new filter are more nearly gaussian, and the frequency responses cross at 3 db of attenuation rather than 6 db. However, since no measurements have yet been made of the spectra of channel signals from the modified channels nothing further shall be said of them here.

Channel signals change slowly enough that it is possible to associate them with the actual syllables to which they correspond. Figure 5 shows some of the channel signals which were analyzed. The sentence is "George is the villain of the novel," by a male speaker. These signals are from channels in different parts of the audio band, and they are typical of the channel signals one observes from other speakers and sentences. It is easy to see that relatively low frequencies prodominate in these channel signals, as there are no very rapid changes in signal amplitude during the two second period shown.

The computation of the power spectrum is accomplished by taking the Fourier transform of the autocorrelation function of the signal. The signal is assumed to be zero except during the two second interval during which it is sampled. Therefore the integrals defining the autocorrelation function and the Fourier transform can be approximated by finite sums of products. Eight thousand registers were set aside for a table of cosines and the spectrum was computed at points 0.125 cps apart in the range of 0-32 cps.* The time required for the computation of the spectrum of a channel signal was 40 seconds. The outputs of the computations were stored on punched paper tape for further processing. Other features of the program permitted displaying the channel signal, the autocorrelation function, and the power spectrum or its square root, and plotting a graph of the spectrum.

*Although the spectral resolution is 0.5 cps, plots of the spectrum are unpleasantly coarse unless points are computed at frequency intervals somewhat less than that.

III. RESULTS AND CONCLUSIONS

The twenty sentences played through eighteen different channels produced 360 channel signals, each of which was analyzed. Figure 6 shows a typical channel signal spectrum, resulting from the sentence, "Where were you when we were all here?," by a male speaker. This is a magnitude spectrum (the square root of the power spectrum), and the scale is linear. Figure 7 shows the spectrum of a channel signal from the same channel and the same sentence but by another male speaker. The considerable difference in fine structure between the two is due to slight differences in the rate of speaking. The fine structure itself is caused by the presence of more than one stressed syllable in most channel signals. Nulls are produced which disguise the average properties of channel signal spectra. Therefore all the spectra of channel signals from a particular channel were averaged* together by the computer. This eliminated most of the peaks and nulls and resulted in a smoother curve. Figure 8 shows such channel spectrum averages for three of the channels. They are averages over twenty sentences. The spectra in Figure 8 are characterized by a slope of -1 db/cps over the range of one to thirty cps. This agrees well with Holmes' results. A sharp drop of about 10 db from the dc value, which does not appear in Holmes' curves, is very prominent, however, in Figure 8.

The computer was also programmed to find, for each channel, the sampling rate necessary to reproduce the channel signals with a mean square

---

* In the spectrum computation the autocorrelation function is normalized with respect to its value at zero. This is equivalent to forcing all the channel signals to have constant energy.

error less than two percent. This sampling rate should be twice the frequency below which 99% of the energy of the channel signal is contained.* Figure 9 shows the results of the computation. The horizontal axis is channel center frequency and is divided into three regions corresponding to the three sets of filters in Table 1. The vertical axis is the 99^{th} percentile frequency - it should be doubled to give the sampling rate required. The average "99% frequency" for a channel is represented by the height of a dot, and the mean deviations of the "99% frequency" for individual channel signals are indicated above and below the channel average. These mean deviations are typically about 2 cps. The conclusions most readily drawn from this graph are that 50 samples per second will suffice for accurate transmission of almost all the channel signals, and that significant reduction of the transmission rate could be made only for the very lowest frequency channels.

This conclusion must be qualified in at least two ways. First, there is no reason to believe that a small mean square error is the proper criterion for transmission of channel signals. It may well be that it is more important to preserve the transitions accurately than to be accurate during the steady portions of the speech. Second, this result ignores the effects of quantization of the samples, which will usually result in a mean square error considerably larger than two percent. In the final analysis only subjective testing of the outputs of otherwise identical vocoders with different sampling rates can suffice to say what sampling rates are necessary for adequate intelligibility, for pleasant quality, or for excellent fidelity.

The odd peak in Fig. 9 in the vicinity of 700 cps has not been explained; it would be interesting to see if further data corroborates this

*K. L Jordan, "Discrete Representations of Random Signals," TR 378, Research Laboratory of Electronics (M. I. T., 14 July 1961)

effect. It would also be interesting to see if Holmes' observation that wider band pass filters produce narrower channel signal spectra is confirmed, and whether this is a general rule or a statistical effect. Another experiment of interest would be to measure at least some channel signal spectra for channels with different kinds of filters; it is suspected that when a pitch harmonic moves in or out of the band of a filter with sharp skirts, the channel signal might vary too rapidly to be sampled adequately at 50 cps. On the other hand, filters with broader skirts than those used in this experiment might not require as high a sampling rate as 50 cps.

TABLE 1

Band Pass Filters Used for the Experiment

Low Frequency Filters (Every Channel)

60 - 180 cps	780 - 900 cps
180 - 300 cps	900 - 1020 cps
300 - 420 cps	1020 - 1140 cps
420 - 540 cps	1140 - 1260 cps
540 - 660 cps	1260 - 1380 cps
660 - 780 cps	

Middle Frequency Filters (Every Third Channel)

1500 - 1620 cps	2580 - 2700 cps
1860 - 1980 cps	2940 - 3060 cps
2220 - 2340 cps	

High Frequency Filters (360 cps Wide)

3300 - 3660 cps	4380 - 4740 cps

FIGURE CAPTIONS

Fig. 1 Vocoder Analyzer Channel

Fig. 2 Experimental Procedure

Fig. 3 Low Pass Filter Frequency Response

Fig. 4 Frequency and Impulse Response of Lerner Filter

Fig. 5 Some Typical Channel Signals

Fig. 6 Channel Signal Spectrum

Fig. 7 Channel Signal Spectrum

Fig. 8 Averaged Spectra of Channel Signals

Fig. 9 Sampling Frequencies Based on Average Spectra

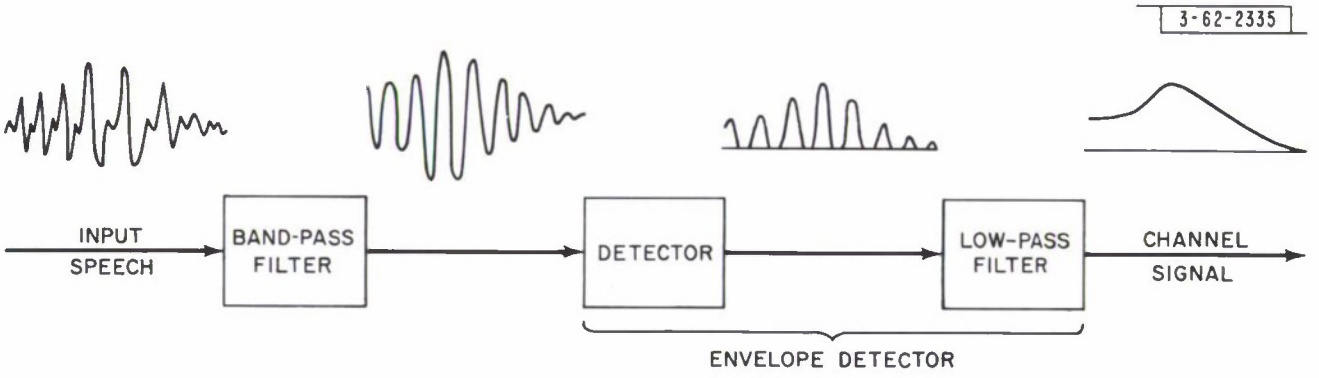

3-62-2335
INPUT
SPEECH
BAND-PASS
FILTER
DETECTOR
LOW-PASS
FILTER
CHANNEL
SIGNAL
ENVELOPE DETECTOR

Fig. 1

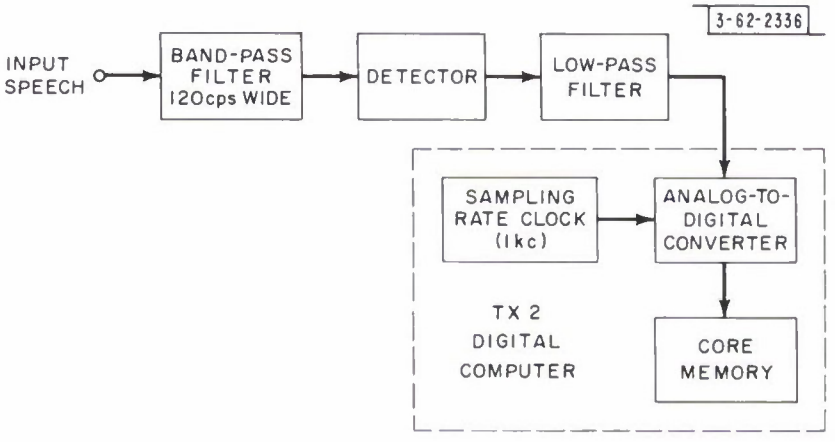

3-62-2336
INPUT SPEECH
BAND-PASS FILTER 120cps WIDE
DETECTOR
LOW-PASS FILTER
SAMPLING RATE CLOCK (1kc)
ANALOG-TO-DIGITAL CONVERTER
TX 2 DIGITAL COMPUTER
CORE MEMORY

Fig. 2

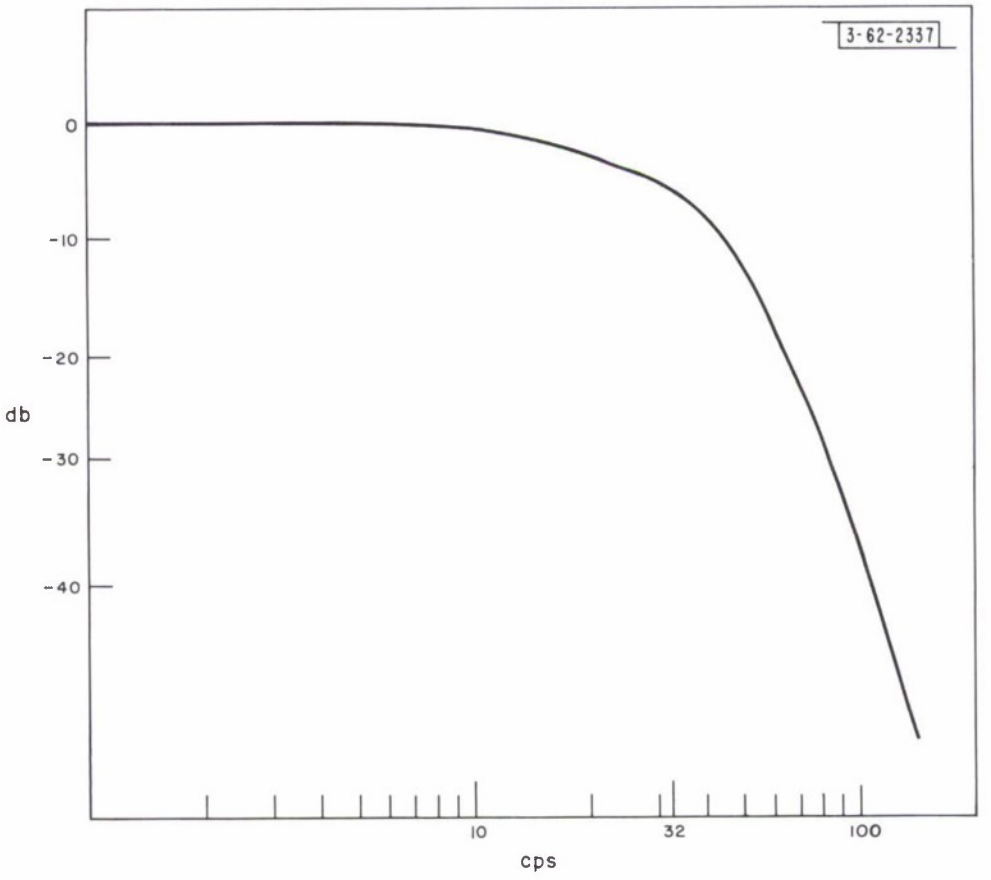
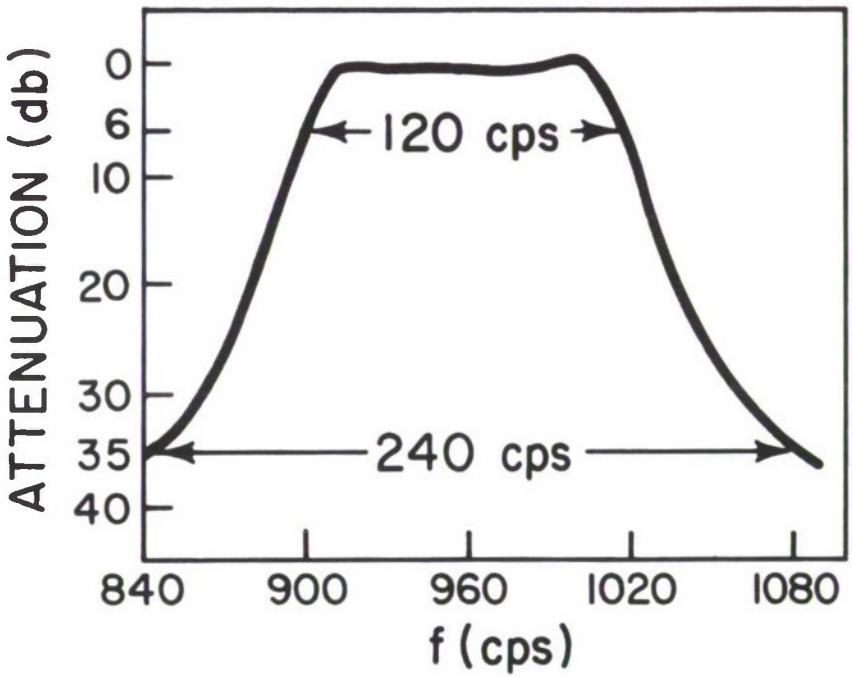
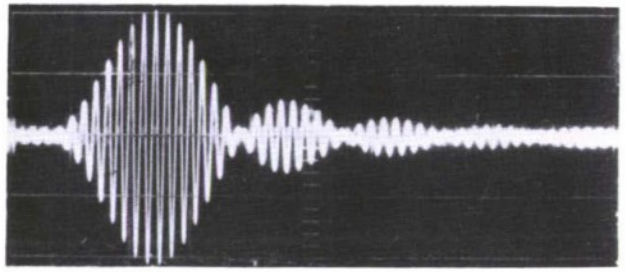


Fig. 3





5 ms/div

C62-219

Fig. 4

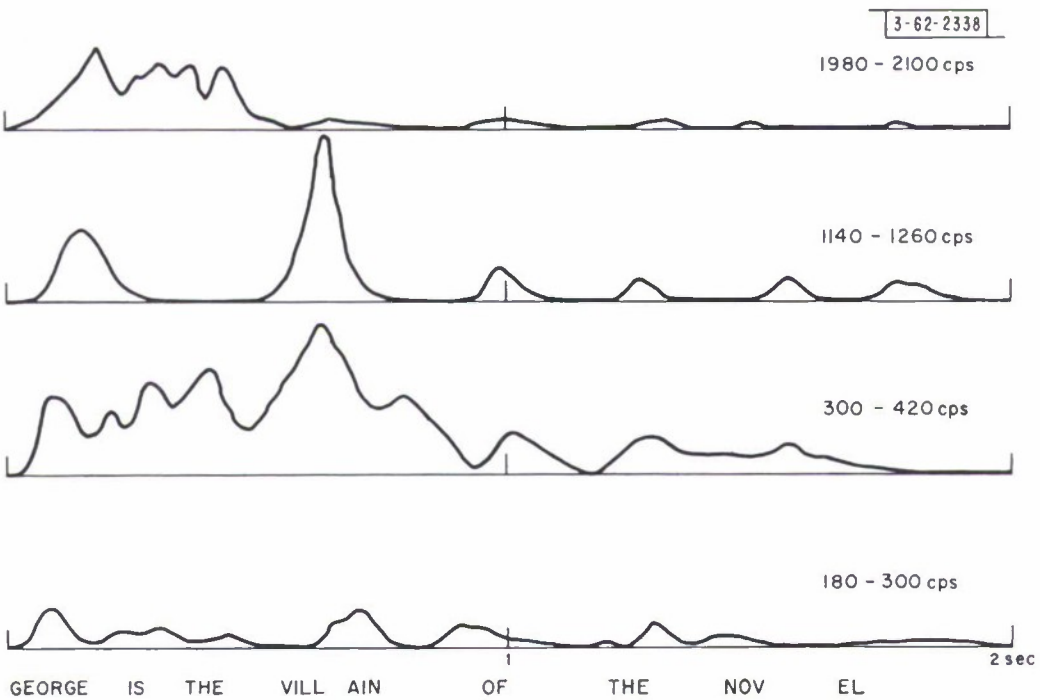

3-62-2338
1980 - 2100 cps
1140 - 1260 cps
300 - 420 cps
180 - 300 cps
1
2 sec
GEORGE IS THE VILL AIN OF THE NOV EL

Fig. 5

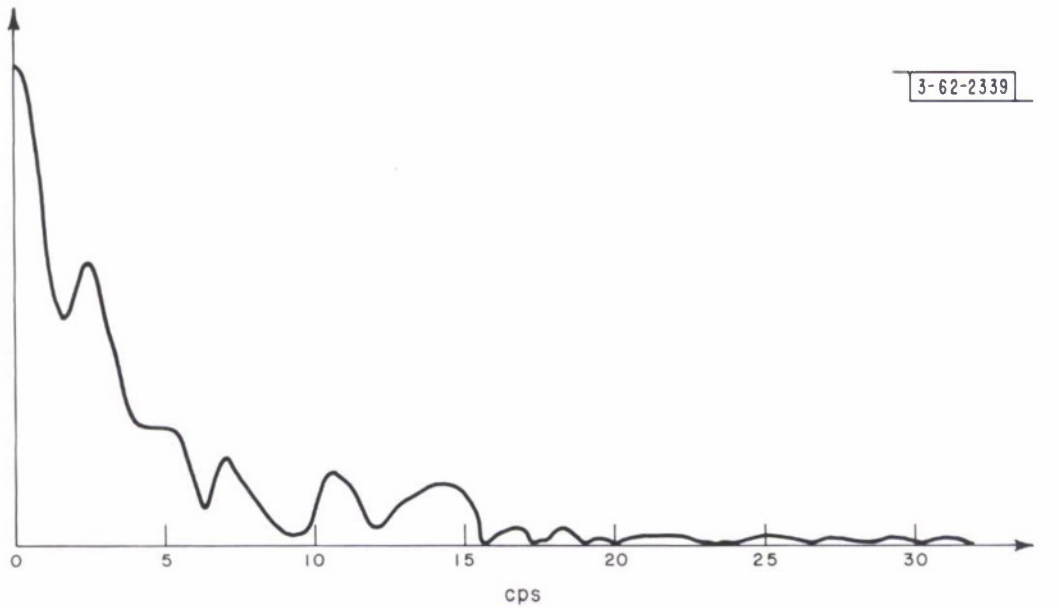


Fig. 6

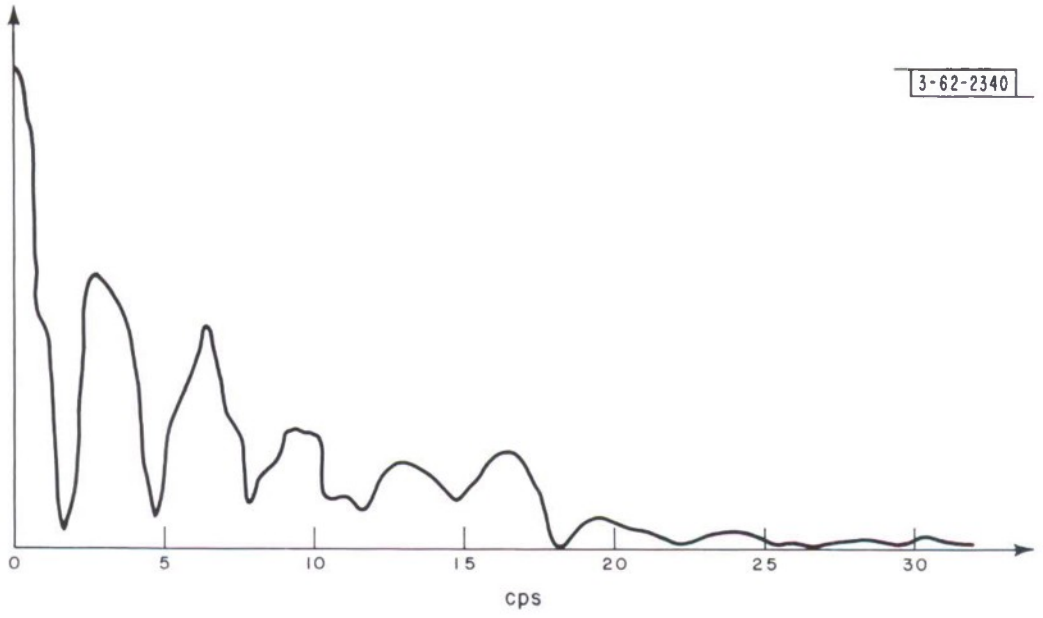


Fig. 7

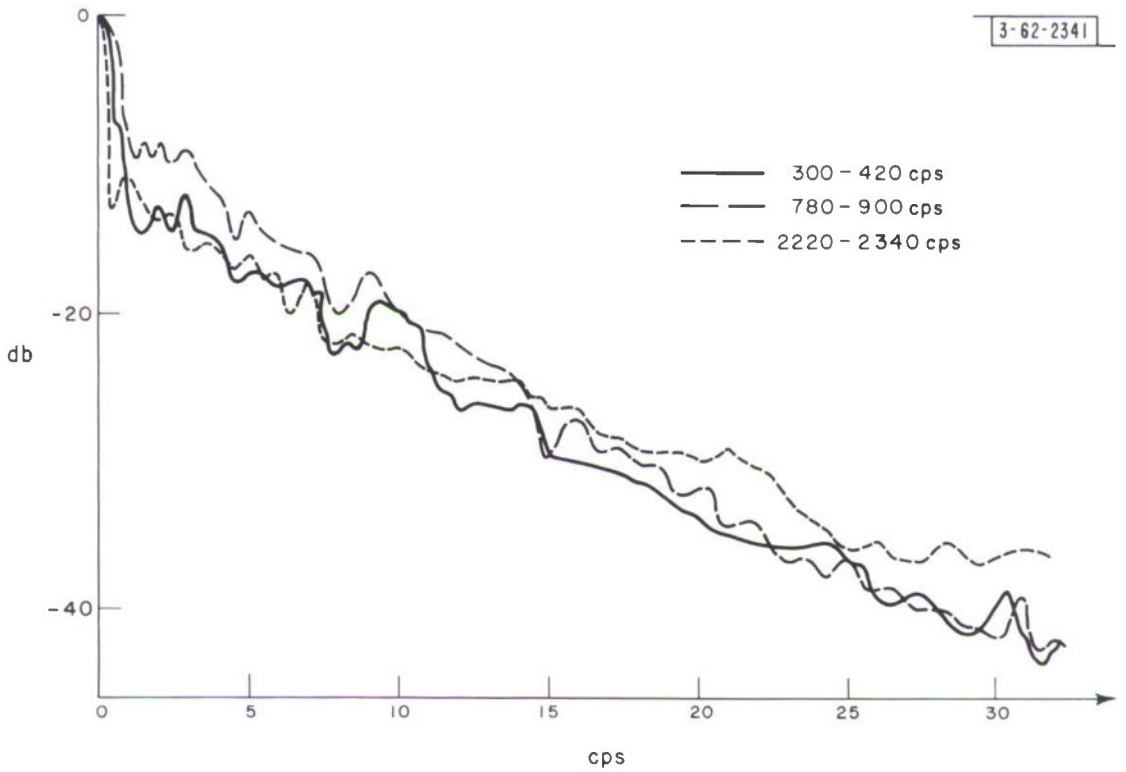

3-62-2341
300 - 420 cps
780 - 900 cps
2220 - 2340 cps
0
-20
-40
db
0
5
10
15
20
25
30
cps

Fig. 8

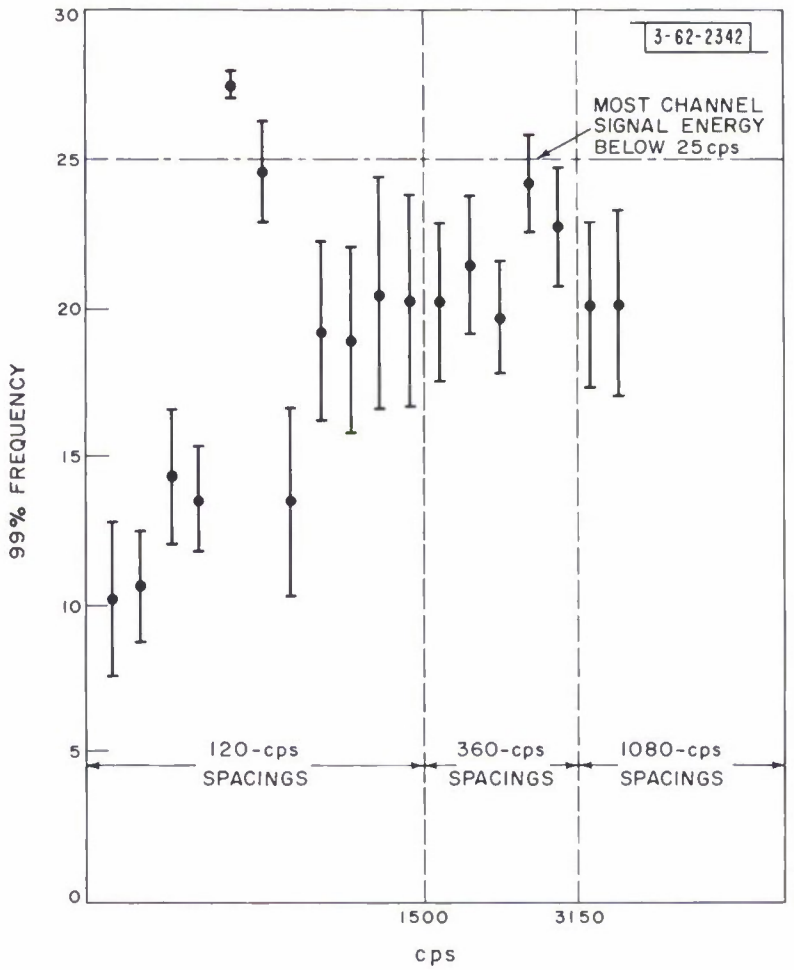

3-62-2342
MOST CHANNEL
SIGNAL ENERGY
BELOW 25 cps
99% FREQUENCY
30
25
20
15
10
5
0
120-cps
SPACINGS
360-cps
SPACINGS
1080-cps
SPACINGS
1500
3150
cps

Fig. 9

DISTRIBUTION LIST

Director's Office

W. B. Davenport, Jr.
W. H. Radford

Division 2

F. C. Frick
C. D. Forgie
J. W. Forgie
M. L. Groves
C. K. McElwain
J. L. Mitchell
A. I. Schulman
O. G. Selfridge
J. E. K. Smith
W. S. Torgerson
M. L. Wendell

Division 3

M. L. Stone

Division 6

G. P. Dinneen
P. E. Green, Jr.
R. M. Lerner
W. E. Morrow, Jr.
R. T. Prosser
L. J. Ricardi
H. Sherman
V. J. Sferrino

AFL

N. Levine

Group 62

E. J. Aho
N. L. Daggett
R. E. Drapeau
J. A. Dumanain
B. Gold
J. N. Harris
D. A. Hunt
B. H. Hutchinson, Jr.
K. L. Jordan
D. Karp
R. S. Kennedy
I. L. Lebow
R. W. MacKnight
A. C. Parker
C. M. Rader (24)
P. Rosen
J. Tierney
N. C. Vlahakis
R. V. Wood
J. Wozencraft
Group 62 Files (5)